INRIA

CENTRE DE RENNES
IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tel (3) 954 90 20

Rapports de Recherche

N° 476

ESTIMATION DE L'ORDRE D'UN PROCESSUS ARMA À L'AIDE DE RÉSULTATS DE PERTURBATIONS DE MATRICES

Jean-Jacques FUCHS

Janvier 1986

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CÉDEX
FRANCE
Téléphone : 99 36 20 00
Télex : UNIRISA 950 473 F
Télécopie : 99 38 38 32

Publication N° 275
Décembre 1985
40 pages

ESTIMATION DE L'ORDRE D'UN PROCESSUS ARMA A L'AIDE DE RESULTATS DE PERTURBATIONS DE MATRICES

Jean-Jacques FUCHS

Résumé.– On propose une méthode d'estimation de l'ordre d'un processus ARMA scalaire. La méthode est basée sur la relation qui existe entre le rang de certaines matrices constantes sur la suite des covariances du processus et l'ordre. On considère une suite de matrices symétriques de dimension croissante et on réalise un test sur la valeur propre de plus petite valeur absolue de chacune de ces matrices pour décider si elle doit être considérée comme étant égale à zéro (et le test arrêté) ou non (et la matrice suivante analysée). La justification du test fait intervenir les propriétés asymptotiques des estimées des covariances des séries temporelles et des résultats de perturbations de matrices.

.../...

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(L. A. 227)
UNIVERSITÉ DE RENNES 1 I. N. S. A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(LABORATOIRE DE RENNES)

ARMA - ORDER ESTIMATION VIA MATRIX PERTURBATION THEORY

Jean-Jacques FUCHS, Member IEEE
IRISA - Université de Rennes I
Campus de Beaulieu - 35042 Rennes Cedex, France

and

GRECO CNRS - Systèmes adaptatifs

Abstract. - Allmost all methods for estimating the model order in system identification -either involve the maximization of likelihood functions, a time-consuming task, -or rely upon the determination of the rank of (covariance) matrices, a task generally achieved by means of heuristic tests. We propose a scheme belonging to this second class of methods for which, however we theoretically justify the test. Using the asymptotic properties of sample serial covariances and some results from matrix perturbation theory, we obtain the statistical distribution of the "smallest" eigenvalue of -say- the Hankel matrix build upon the estimated covariances, under the hypothesis that the corresponding exact Hankel matrix possesses one single zero eigenvalue. This allows us to develop and justify a test which moreover only requires the knowledge of the "smallest" eigenvalue and an associated eigenvector. A complete eigen-decomposition is thus not necessary further limiting the computations. The new order determination scheme is compared on simulated examples to the more time consuming approaches based on likelihood maximizations, the performance appear to be comparable.

PAPIER RÉCUPÉRÉ ET RECYCLÉ

PI 269 **Modélisation et gestion d'univers 3D (une première approche à partir du logiciel PADL-2**
Gérard Hégron, Patrick Rives – 60 pages ; Octobre 85.

PI 270 **LGRC : Un langage pour la synthèse d'images par lancer de rayons**
Bruno Arnaldi, Kadi Bouatouch, Thierry Priol – 30 pages ; Novembre 85.

PI 271 **Towards specification and proof of asynchronous systems**
Boubakar Gamatié – 30 pages ; Novembre 85.

PI 272 **Architecture pour les opérations geométriques en synthèse d'images par facettes**
François Charot, Franck Rousée – 24 pages ; Novembre 85.

PI 273 **Madmacs : a new VLSI layout macro editor**
Patrice Frison, Eric Gautrin – 12 pages ; Novembre 85.

PI 274 **Détection de pannes et reconfiguration automatique**
Michèle Basseville – 26 pages ; Novembre 85.

PI 275 **Estimation de l'ordre d'un processus Arma à l'aide de résultats de perturbations de matrices**
Jean-Jacques Fuchs – 40 pages ; Décembre 85.

PI 276 **Detection and diagnosis of changes in the eigenstructure of nonstationary multivariable systems**
Michèle Basseville, Albert Benveniste, Georges Moustakides, Anne Rougée – 44 pages ; Décembre 85.

Jean-Jacques FUCHS

ESTIMATION DE L'ORDRE D'UN PROCESSUS ARMA A L'AIDE DE RESULTATS DE PERTURBATIONS DE MATRICES

Publication interne
n° 275

Décembre 1985

1 - INTRODUCTION

Because the maximization of the likelihood function does not allow to determine structural parameters such as the orders in autoregressive moving average (ARMA) processes, other criteria have been proposed by Akaike [1,2], Schwarz [3] and Rissanen [4,5]. All these criteria, though derived using different approaches, are not essentially different when applied to scalar ARMA-processes and require the calculation of a maximum likelihood estimate for each possible order-pairs. While these methods are attractive in treating AR-processes, their use in ARMA processes is limited. This is due to the complexity in the evaluation of the likelihood functions and the lack of consistency and stability in most ARMA identification procedure when the orders tested are higher than the true orders. Different solutions [6,7] have been proposed to overcome these difficulties and computationally attractive algorithms developped.

In this paper, we present a fully different approach, which belongs to the order estimation methods based on the determination of the rank of matrices. It is well known that the orders of an ARMA process can be obtained by investigation of the rank of covariance matrices [8,9] with non-symmetric Toeplitz structure or the rank of Hankel matrices build upon the covariance sequence of the process. Make a decision about the rank of a matrix is a difficult task and more so if the matrix is contaminated with errors, which is always the case in our context since the matrix is based on estimated covariances. The most commonly used methods for evaluating the rank of a matrix rely upon the

singular value decomposition, the QR factorization, the inverse cross-product matrix [10]. These approaches have been applied to the estimation of the orders of ARMA processes [11,9,12], they are however only tools for helping make a decision and do by no mean yield a test. Further analysis is required to be able to decide, say, how many singular values are "zero", i.e. to be able to justify a threshold against which to check the singular values. To our knowledge no such analysis has been proposed in the litterature.

In this paper, we present such an analysis. The order determination scheme we propose, consists in testing, for a sequence of symmetric (Hankel) matrices, their smallest singular value against a theoretically justified threshold to decide whether it should be declared zero or not and thus the test terminated or the next matrix in the sequence considered. (Note that for a symmetrix matrix, the singular values are the absolute values of the eigenvalues). The threshold we propose is obtained using the asymptotic properties of the estimated covariances of stationary time series [13] and some results from matrix perturbation theory [14]. The test is thus only justified asymptotically in the number of observations and in order to verify its usefulness some simulation results are proposed. They seem to indicate that the theoretical predictions are valid for quite small sample sizes and that the proposed order estimation scheme has performances similar to those of the more time consuming approaches based on likelihood maximizations.

2 - PROBLEM STATEMENT

Let y_t be a scalar autoregressive moving average (ARMA) process of order (p,q) :

$$y_t + a_1 y_{t-1} + \ldots + a_p y_{t-p} = e_t + c_1 e_{t-1} + \ldots + c_q e_{t-q} \qquad (1)$$

where e_t is a gaussian white noise sequence with zero mean and variance σ_e^2 and the roots of the two coprime polynomials

$$A(z) = 1 + a_1 z + \ldots + a_p z^p \qquad a_p \neq 0$$

$$C(z) = 1 + a_1 z + \ldots + c_q z^q \qquad c_q \neq 0$$

are strictly outside the unit circle.

We shall define a procedure which, using a sequence of length T of the outputs $\{y_t\}$ allows to estimate p and q. In a first step we obtain an estimate of :

$$n = \max(p,q) \qquad (2)$$

the order of the stochastic process $\{y_t\}$.

In section 3, we define the test strategy we propose and justify it under exact covariance information. We then introduce some matrix perturbation results in section 4. These results are used in section 5 to establish the statistical properties allowing to build a test. In section 6, the actual test procedure is defined. A simplified way to perform it is proposed in section 7 together with some simulation results.

Let us specify the following notations and symbols to be used in the paper : ε is a small positive real number ($\varepsilon \ll 1$) which constitutes the basis of the different orders of magnitudes

$f(\varepsilon) = o(\varepsilon^n)$ implies that $f(\varepsilon)/\varepsilon^n \to 0$ as $\varepsilon \to 0$

$f(\varepsilon) = O(\varepsilon^n)$ implies that $f(\varepsilon)/\varepsilon^n$ is bounded as $\varepsilon \to 0$.

We will also use these notations with $n = 0$, and thus write $f(\varepsilon) = O(1)$ to indicate that $\varepsilon \ll f(\varepsilon) \ll 1/\varepsilon$. When applied to a matrix, the implication are to be considered component-wise.

3 - THE TEST STRATEGY

Let $\{r_o, r_1, \ldots\}$ be the covariance sequence of the process, i.e. :

$$r_k = E(y_t\, y_{t-k}) \qquad r_k = r_{-k}$$

and let $\{\hat{r}_o, \hat{r}_1, \ldots\}$ be the estimated sample covariances using T observations :

$$\hat{r}_k = \frac{1}{T-k} \sum_{t=k+1}^{T} y_t\, y_{t-k} \qquad k \geqslant 0 \ . \tag{3}$$

We shall express $\hat{r}_k$ as :

$$\hat{r}_k = r_k + \delta r_k \tag{4}$$

where since the estimates (3) are consistent and asymptotically normal [15,13], the dominant term of δr_k is of order $T^{-1/2}$ -(denoted $O(T^{-1/2})$)-, i.e. the variance of δr_k is of order 1/T.

Introduce the following covariance matrix with Toeplitz structure :

$$R(\hat{p},\hat{q}) = E(\phi(t,\hat{p},\hat{q})\ \psi(t,\hat{p})^T)$$

$$= \begin{bmatrix} r_{\hat{q}} & r_{\hat{q}-1} & \cdots & r_{\hat{q}-\hat{p}+1} \\ r_{\hat{q}+1} & & & \\ \vdots & & & \\ r_{\hat{q}+\hat{p}-1} & & & r_{\hat{q}} \end{bmatrix} \tag{5}$$

where $\hat{p}$ and $\hat{q}$ are positive integers and

$$\phi(t,\hat{p},\hat{q}) = [y_{t-\hat{q}-1} \cdots y_{t-\hat{q}-\hat{p}}]^T$$
$$\psi(t,\hat{p}) = [y_{t-1}\ y_{t-2} \cdots y_{t-\hat{p}}]^T \qquad (6)$$

are two $\hat{p}$-dimensional vectors. The use of this matrix for testing the orders of (1) was suggested in [11,9,12] and its rank properties are well known [8,9]. Define

$$n^* = \min(\hat{p}-p,\ \hat{q}-q) \qquad (7)$$

and $\rho(\hat{p},\hat{q})$ = determinant $R(\hat{p},\hat{q})$, then

$$\rho(\hat{p},\hat{q}) = 0 \qquad \text{for} \quad n^* > 0$$
and $$\qquad (8)$$
$$\rho(\hat{p},\hat{q}) \neq 0 \qquad \text{for} \quad n^* = 0 \ .$$

As indicated in [9] no general statement can be made for $n^* < 0$, but the matrix is generically non-singular.

Rather than to use $R(\hat{p},\hat{q})$ we shall consider :

$$H(\hat{p},\hat{q}) = R(\hat{p},\hat{q})J \qquad (9)$$

where :

$$J = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}$$

permutes the columns of $R(\cdot)$. Since J is a nonsingular matrix, $H(\cdot)$ and $R(\cdot)$ share the same rank properties. $H(\hat{p},\hat{q})$ however is now a symmetric (Hankel) matrix. In many instances it is preferable to work with symmetric matrices ; in our case, in a matrix-perturbation context, symmetric matrices are generally better conditionned [19].

Note also that :

$$H(\hat{p},\hat{p}) = \begin{bmatrix} r_1 & r_2 & \cdot & \cdot & r_{\hat{p}} \\ r_2 & & & & \\ \cdot & & & & \\ \cdot & & & & \\ r_{\hat{p}} & & & & r_{2\hat{p}-1} \end{bmatrix} \qquad (10)$$

is the "standard" Hankel matrix associated with the stochastic process which is known to be of rank $n = \max(p,q)$ for $\hat{p} \geq n$.

Under exact covariance information, there are many different ways to determine the orders p and q. We shall consider the following test strategy which is based on and easily justified using (7,8).

TEST STRATEGY

Step 1 : Test for singularity $H(\hat{p},\hat{p})$ for increasing $\hat{p} = 1,2,...$ and stop for $\hat{p} = j$ with $H(j,j)$ the first singular matrix in the sequence. Then set $n = j-1$ (= max (p,q)) and go to step 2.

Step 2 : Test for non-singularity $H(n+1, n-i)$ for $i = 0,1,2,...$. If $H(n+1, n)$ is non-singular, go to step 3, else stop for $n-i = n-j$ with $H(n+1, n-j)$ the first non-singular matrix. The process is ARMA $(n, n-j)$. End.

Step 3 : Test for non-singularity $H(n-i, n+1)$ for $i = 0,1,2,...$ and stop for $n-i = n-j$ with $H(n-j, n+1)$ the first more singular matrix. The process is ARMA $(n-j, n)$. End.

Since the true covariances are unknown, this strategy will be applied to $\hat{H}(\hat{p},\hat{q})$ which is defined similarly to $H(\hat{p},\hat{q})$ with the sample covariances replacing the true covariances. As in (4) we express $\hat{H}(\hat{p},\hat{q})$ as :

$$\hat{H}(\hat{p},\hat{q}) = H(\hat{p},\hat{q}) + \delta H(\hat{p},\hat{q}) \qquad (11)$$

where the dominant term of $\delta H(\cdot)$ is of order $T^{-1/2}$. The rank properties of $H(\cdot)$ given above then hold for $\hat{H}(\cdot)$ only asymptotically in T and for finite sample size $\hat{H}(\hat{p},\hat{q})$ is generically a full rank $(\hat{p},\hat{p})$-matrix. In order to obtain estimates of p and q when applying the above strategy to $\hat{H}(\cdot)$, we need a test for singularity of matrices. To our knowledge, no such test has been proposed and justified in the litterature.

Note that in the proposed test-strategy, the matrices tested for singularity admit at most one zero eigenvalue. Thus letting $\hat{\lambda}_{min}(\hat{p},\hat{q})$ denote the smallest, in absolute value, eigenvalue of the estimated symmetric matrix $\hat{H}(\hat{p},\hat{q})$:

$$\hat{\lambda}_{min}(\hat{p},\hat{q}) = \lambda_j\ (\hat{H}(\hat{p},\hat{q}))$$

where $\lambda_i(\hat{H}(\cdot))$ denotes the different eigenvalues of $\hat{H}(\cdot)$ and :

$$j = \arg \min_{i\in[1,\hat{p}]} |\lambda_i(\hat{H}(\hat{p},\hat{q}))| \quad ,$$

it will be enough to be able to test the hypothesis $\hat{\lambda}_{min}(\hat{p},\hat{q}) = 0$. In the next two sections, we shall indeed derive the statistical properties of $\hat{\lambda}_{min}(\hat{p},\hat{q})$ for matrices $\hat{H}(\hat{p},\hat{q})$ which possess a "zero" eigenvalue, i.e. for which (7) $n^* = \min(\hat{p}-p,\ \hat{q}-q) = 1$. At this point, one can expect these properties to depend upon T the sample size, the process itself and σ_e^2 the driving noise power since no normalization has been performed on the data.

4 - MATRIX PERTURBATION RESULTS

Let us turn our attention to some results from matrix perturbation theory, that is to theorems which give an estimate of how much the eigenvalues and eigenvectors of the matrix $A+\varepsilon B$ can differ from those of A for small ε. Applying these results to relation (11) :

$$\hat{H}(\cdot) = H(\cdot) + \delta H(\cdot)$$

where $\delta H(\cdot)$ is of order $T^{-1/2}$, will allow us to obtain some deterministic relations between corresponding eigen-elements of $\hat{H}(\cdot)$ and $H(\cdot)$. It appears that in our context, given the information available upon $\delta H(\cdot)$, there is only one of these relations which can be used to gain some precise knowledge upon the discrepancy between corresponding eigen-elements. This relation relates $\hat{\lambda}_{min}(\hat{p},\hat{q})$ for matrices with $n^* = 1$ to its corresponding zero eigenvalue of the exact matrix.

In our test strategy, the matrices for which $n^* = 1$, belong to one of the following two classes :

class 1 : $H(\hat{p},\hat{q})$: $\hat{p} = p+1$; $\hat{q} = q+i$, $i = 1,2,\ldots p-q+1$

class 2 : $H(\hat{p},\hat{q})$: $\hat{p} = p+i$, $i = 1,2,\ldots,q-p+1$; $\hat{q} = q+1$

with p, q the true ARMA orders.

It is an easy task to verify that class-1 matrices, which are of constant dimension (p+1), admit as an eigenvector associated with the single zero eigenvalue :

$$X^T = [a_p \; a_{p-1} \; \ldots\ldots \; a_1 \; 1] \qquad (12)$$

where the a_i's are the AR-coefficients of the ARMA-process (1). The eigen-

vector X is the solution of the linear system of equations $H(p+1, q+i)X = 0$, which can be rewritten $R(p+1, q+i)JX = 0$ where one recognizes (p+1) consecutive Yule-Walker equations. An analoguous reasoning applied to class-2 matrices of varying dimension (p+i), yields the following eigenvector :

$$X^T = [\underbrace{0 \ \dots \ 0}_{i-1 \text{ zeros}} \ a_p \ a_{p-1} \ \dots \ 1]$$

We shall only consider class-1 matrices in the sequel, however the analysis to be performed is readily transposable to class-2 matrices, and the result will be expressed in terms valid for both classes.

The pertubation results we need are quite elementary and proofs can be found in [14,16]. These are however quite involved. We propose in Appendix A a simple, self-contained proof derived from [17]. Applied to the unique zero eigenvalue of a class-1 matrix, they state that the corresponding perturbed eigenvalue $\hat{\lambda}_{min}(p+1, q+i)$ verifies (A5) :

$$\hat{\lambda}_{min}(p+1, q+i) = \bar{X}^T \, \delta H(p+1, q+i) \bar{X} + O(T^{-1}) \qquad (13)$$

where $\bar{X} = X/\|X\|$ is the normalized true eigenvector (12) ; while $\hat{\bar{X}}$ the normalized perturbed eigenvector satisties (A9) :

$$\hat{\bar{X}} = \bar{X} + O(T^{-1/2}) \qquad (14)$$

These results (13,14) hold under the restriction that the non-zero eigenvalues of $H(p+1, q+i)$ are of an order of magnitude greater than $O(T^{-1/2})$. For a given process or a given class-1 Hankel matrix, this restriction is always satisfied for large enough sample sizes T. This kind of limitation is not unexpected and is common to all order-determination tests, i.e. all tests rely upon asymptotic properties and are thus only valid -or better- justified for large T.

with :

$$v_t = \bar{X}^T \phi(t, p+1, q+i)$$

$$w_t = \psi^T(t, p+1) J \bar{X}$$

and hence from (6) and (12) :

$$\begin{aligned} v_{t+q+i+1} &= (1/\|X\|)(y_{t-p} + a_1 y_{t-p-1} + \ldots + a_p y_t) \\ w_t &= (1/\|X\|)(y_{t-1} + a_1 y_{t-2} + \ldots + a_p y_{t-p-1}) \end{aligned} \tag{20}$$

In (19) $\hat{\lambda}_{min}(\cdot)$, seen as a random variable, appears as the empirical cross-covariance of lag zero of two stochastic processes v_t and w_t which are linear functions (20) of the observed process y_t. These two processes will be further identified as stationary moving average processes and $\hat{\lambda}_{min}(\cdot)$ will thus satisfy a central limit theorem, i.e. asymptotically in the sample size T, $\hat{\lambda}_{min}(\cdot)$ will be a gaussian random variable with a variance of order T^{-1}.

There are two different ways to investigate the properties of the processes v_t and w_t, the "geometric" and the "algebraic". We shall use both of them. From (1) and (20) one verifies that w_t is a moving average process of order q :

$$w_t = (1/\|X\|)\ (e_{t-1} + c_1 e_{t-2} + \ldots + c_q e_{t-q-1}) \ . \tag{21}$$

The process v_t would appear as the AR-part of the reverse ARMA-process of ARMA process $y_t(1)$ and is thus a backward moving-average process of order q driven by -say b_t- the backward innovation process of y_t :

$$v_{t+p+i+1} = (1/\|X\|)\ (b_{t-q} + c_1 b_{t-q+1} + \ldots + c_q b_t) \ . \tag{22}$$

The two processes v_t and w_t have the same autovariance function because the reverse process of an ARMA-model follows the same ARMA-model. These facts follow also easily from (19) in an algebraic way. Define

$$\gamma_k = E(v_{t-k}\, v_t)$$

then, with simplified notations :

$$\begin{aligned}\gamma_k &= E(\bar{X}^T\, \phi(t-k,\cdot)\, \phi^T(t,\cdot)\bar{X}) \\ &= \bar{X}^T\, E(\phi(t-k,\cdot)\, \phi^T(t,\cdot))\bar{X} \\ &= \bar{X}^T\, R(p+1,\, k)\bar{X} \quad , \end{aligned} \tag{24}$$

consider now :

$$\begin{aligned}E(w_t\, w_{t-k}) &= E(\bar{X}^T\, J^T\psi(t,\cdot)\, \psi^T(t-k,\cdot)J\bar{X}) \\ &= \bar{X}^T\, J^T\, R^T(p+1,\, k)J\bar{X}\end{aligned}$$

which, since for Toeplitz matrices :

$$R(p+1,k) = J^T\, R^T(p+1,\, k)J$$

establishes that $E(v_t\, v_{t-k}) = E(w_t\, w_{t-k})$. By definition (9) of $H(\cdot)$, (24) is equivalent to :

$$\gamma_k = \bar{X}^T\, H(p+1,\, k)J\bar{X} \tag{25}$$

which, since $\bar{X}$ (12) satisfies

$$\bar{X}^T\, H(p+1,\, q+i) = 0 \qquad i = 1,2,\ldots, \tag{26}$$

implies

$$\gamma_{q+i} = 0 \qquad i = 1,2,\ldots, \tag{27}$$

a property which characterizes moving average processes of order q. Let us

Since $\bar{X}$ is the true zero eigenvector, combining (11) and (13) yields :

$$\hat{\lambda}_{min}(p+1, q+i) = \bar{X}^T \hat{H}(p+1, q+i)\bar{X} + O(T^{-1}) \qquad (15)$$

This purely deterministic relations says that if the unperturbed normalized eigenvector $\bar{X}$ of $H(\cdot)$ is known, a good estimation of $\hat{\lambda}_{min}(\cdot)$, the perturbed zero-eigenvalue, is given by $\bar{X}^T\hat{H}(p+1, q+i)\bar{X}$, an non-surprising result if one remembers that indeed one has identically :

$$\hat{\lambda}_{min}(p+1, q+i) = \hat{\bar{X}}^T \hat{H}(p+1, q+i)\hat{\bar{X}} \qquad (16)$$

with $\hat{\bar{X}}$ (14) the normalized perturbed eigenvector. Note, that unless a great amount is known about $\delta H(\cdot)$, perturbation results only give quite poor information such as the order of magnitude (see (14)). The situation is somehow more favorable for zero eigenvalues but relation (15) is still useless in general since $\bar{X}$ is unknown and replacing it by estimate $\hat{\bar{X}}$ yields the trivial relation (16).

5 - STATISTICAL PROPERTIES OF $\hat{\lambda}_{MIN}$

In order to analyse the statistical properties of $\hat{\lambda}_{min}(\cdot)$ seen as a sample of a random variable, we consider the deterministic relation (15) in a different way. In the sequel, we shall neglect higher order terms and thus replace equality signs by $\cong$. Relation (15) is thus re-written :

$$\hat{\lambda}_{min}(p+1,\ q+i) \cong \bar{X}^T\ \hat{H}(p+1,\ q+i)\bar{X} \quad . \tag{17}$$

It can be seen as defining the random variable $\hat{\lambda}_{min}(\cdot)$ as a function of the random matrix $\hat{H}(\cdot)$ which by definition (5,9) is equal to :

$$\hat{H}(p+1,\ q+i) = \begin{bmatrix} \hat{r}_{q+i-p} & \hat{r}_{q+i-p+1} & \cdots & \hat{r}_{q+i} \\ \hat{r}_{q+i-p+1} & & & \\ \vdots & & & \\ \hat{r}_{q+i} & & & \hat{r}_{q+i+p} \end{bmatrix} .$$

Or, using (3) and (6) :

$$\hat{H}(p+1,\ q+i) \cong \frac{1}{T} \sum_{t=1}^{T} \phi(t,\ p+1,\ q+i)\psi^T(t,\ p+1)J \tag{18}$$

where we neglect the higher order terms due to the fact that the matrix on the right is non-symmetric.

Combining now (17) and (18) one has :

$$\hat{\lambda}_{min}(p+1,\ q+i) \cong \frac{1}{T} \sum_{t=1}^{T} v_t\ w_t \tag{19}$$

now establish some properties of the cross-covariance sequence of v_t and w_t. These could easily be deduced -geometrically- from the orthogonality properties between the forward MA-process w_t and the backward MA-process v_t, but we prefer to derive them algebraically from (19) and (26) :

$$E(v_t\ w_{t+k}) = \bar{X}^T\ E(\phi(t,\cdot)\ \psi^T(t+k,\cdot))J\bar{X}$$

$$= \bar{X}^T\ H(p+1,\ q+i+k)\bar{X}$$

which, since for class-1 Hankel matrices i is a positive integer and using (26), leads to :

$$E(v_t\ w_{t+k}) = 0 \qquad\qquad k = 1-i,\ 2-i,\ldots,0,1,2,\ldots \qquad (28)$$

which implies -e.g.- that $\hat{\lambda}_{min}(p+1,\ q+i)$ -an estimator of $E(v_t\ w_t)$- is zero mean. Let us compute its variance :

$$E(\hat{\lambda}^2_{min}(\cdot)) \cong E((\frac{1}{T}\ \sum_1^T v_t\ w_t)^2)$$

$$\cong \frac{1}{T^2}\ E((v_1\ w_1 + v_2\ w_2 + \ldots)^2)$$

which, using notation (23) and properties (27,28) gives :

$$E(\hat{\lambda}^2_{min}(p+1,\ q+i)) \cong \frac{1}{T}\ \{\gamma_o^2 + 2\ \sum_1^q\ \gamma_i^2\} \qquad (29)$$

From the asymptotic normality of sample cross-covariances of stationary time-series [13] it thus follows that asymptotically in T, $\hat{\lambda}_{min}(p+1,\ q+i)$ will be a zero-mean gaussian variable with variance of order 1/T given by (29).

Now in order to build a test and to define a threshold we need to evaluate the asymptotic variance (29) from the data. To evaluate the covariances γ_i appearing in (29) and defined in (24,25) we propose to replace the unknown quantities $\bar{X}$, $R(\cdot)$ and $H(\cdot)$ in (24,25) by their estimates $\hat{\bar{X}}$, $\hat{R}(\cdot)$ and $\hat{H}(\cdot)$. The so-introduced errors will only affect higher order terms in the value of the variance (see (11,14)).

The true MA-order which appears in the summation in (29) is also unknown. Since the standing hypothesis is that the tested matrix belongs to class-1 -i.e. is of the form $\hat{H}(\hat{p},\hat{q})$ with $\hat{p} = p+i$ and $\hat{q} = q+i$, $i = 1,2,..$- we implicitly assume that the true MA order is smaller or equal to $\hat{q}-1$. We propose thus to replace q in (29) by $\hat{q}-1$, the terms in excess, if any, being again neglectible (27). The expression allowing to compute the variance, to be denoted $\hat{\sigma}^2_{min}(\hat{p},\hat{q})$, of $\hat{\lambda}_{min}(\cdot)$ from the data is thus :

$$\hat{\sigma}^2_{min}(\hat{p},\hat{q}) \cong \frac{1}{T}\left(\hat{\gamma}_o^2 + 2\sum_{k=1}^{\hat{q}-1} \hat{\gamma}_k^2\right) \qquad (30)$$

with

$$\hat{\gamma}_k = \hat{\bar{X}}^T \hat{R}(\hat{p},k)\hat{\bar{X}} \quad .$$

While this expression is valid and has been derived for class-1 matrices only, it is also applicable to class-2 matrices with $\hat{\bar{X}}$ the normalized eigenvector associated with $\hat{\lambda}_{min}(\hat{p},\hat{q})$.

COMMENTS : *As expected, the properties of* $\hat{\lambda}_{min}(\cdot)$ *depend upon T, the process-model and the driving-noise power.*

For pure AR-processes, v_t *and* w_t *in (19) are respectively the backward and forward innovation processes of* y_t*, i.e. white noises with iden-*

tical variance γ_o *(24). A closer look at this specific case, leads to the following simplified expression of the variance of* $\hat{\lambda}_{min}$ *:*

$$\hat{\sigma}^2_{min}(p+1,\ i) \cong \frac{1}{T}\ \frac{\sigma_e^4}{(X^TX)^2} \qquad i = 1,2,\ldots \qquad (31)$$

where X (12) is the vector of the AR-parameters and σ_e^2 *is (1) the driving noise variance which satisfies* $\sigma_e^2 = X^T R(p+1,\ 0)X = \gamma_o\ X^TX$.

Instead of evaluating the variance (29) of $\hat{\lambda}_{min}(\cdot)$ *from the data as in (30), it is possible to derive an upperbound of it. For* $\{\gamma_i\}$, *the covariance sequence of a MA(q)-process, the following inequality holds [18] :*

$$\sum_{k=1}^{q} \left|\frac{\gamma_i}{\gamma_o}\right| \leqslant \frac{q}{2}$$

and hence from (29)

$$E(\lambda^2_{min}(\cdot)) \leqslant \frac{q+1}{T}\ \gamma_o^2 \qquad (32)$$

The variance γ_o *is equal to (24)*

$$\gamma_o = \bar{X}^T R(p+1,0)\bar{X}$$

and can be bounded by $\lambda_{max}(R(p+1,\ 0))$ *the greatest eigenvalue of the positive definite Toeplitz matrix R(p+1, 0) (5) :*

$$\gamma_o \leqslant \lambda_{max}(R(p+1,\ 0))$$

$$\leqslant \max_{k=0,\ldots,p} t_k \qquad (\leqslant (p+1)r_o)$$

with :

$$t_o = \sum_{0}^{p} |r_i|$$

and

$$t_k = t_{k-1} + |r_k| - |r_{p+1-k}|$$

where the majorant (max t_k) *for* $\lambda_{max}(\cdot)$ *follows from Gerschgorin circle theorem* [19].

6 - THE TEST PROCEDURE

We already indicated the test strategy we propose. Other strategies can be used and developed along the analysis we performed. Note however that in these strategies, as in the one we propose, the matrices to be tested for singularity or non-singularity must admit at most one "zero" eigenvalue since we obtained the statistical properties of $\hat{\lambda}_{min}(\cdot)$ under this restriction.

Within the test strategy described in Section 3, we propose to use the following chi-squared test, to decide whether a given matrix $\hat{H}(\hat{p},\hat{q})$ should be declared singular or not. We have established that if $\hat{H}(\hat{p},\hat{q})$ belongs to Class 1 or Class 2, then $\hat{\lambda}_{min}(\hat{p},\hat{q})$ is asymptotically gaussian with zero mean and variance $\hat{\sigma}^2_{min}(\hat{p},\hat{q})$ of order T^{-1} given in (30), the quantity

$$\mu = \frac{\hat{\lambda}^2_{min}(\hat{p},\hat{q})}{\hat{\sigma}^2_{min}(\hat{p},\hat{q})} \tag{33}$$

is then a chi-squared random variable with 1 degree of freedom. If, on the other hand, $\hat{H}(\hat{p},\hat{q})$ is not "singular", i.e. does not belong to class 1 or 2, then the statistical properties of $\hat{\lambda}_{min}(\hat{p},\hat{q})$ are different and we will consider a rule of the form :

$$\begin{aligned} \mu \leqslant t \quad &\Rightarrow \quad \hat{H}(\hat{p},\hat{q}) \text{ singular} \\ \mu > t \quad &\Rightarrow \quad \hat{H}(\hat{p},\hat{q}) \text{ non-singular} \end{aligned} \tag{34}$$

With the aid of a chi-squared table, one can evaluate the probability of false alarm (declaring $\hat{H}$ non-singular while it is singular) as a function of the threshold t. The so-obtained confidence interval is however valid only asymptotically in T.

The value to be given to the threshold t has to chosen with care, since we are always working with finite sample sizes T. Too high a threshold will lead to declare "singular" matrices that are not. One could indeed establish that if $\hat{H}(\cdot)$ is non-singular $\hat{\lambda}_{min}(\cdot)$ is asymptotically gaussian with non-zero mean $\lambda_{min}(\cdot)$, the corresponding unperturbed eigenvalue and variance of order T^{-1}. Thus, if for the process under investigation $\lambda_{min}(\cdot)$ is small and moreover the sample size T moderate, the quantity μ in (33) ($\mu \cong \hat{\lambda}^2_{min}.T/\text{constant}$) might be quite small and the matrix wrongly declared singular.

Another observation can be made at this point. Remember that, while the statistical properties of $\hat{\lambda}_{min}(\cdot)$ are obtained by mean of a central limit theorem and thus hold only asymptotically in T, the deterministic matrix perturbation result is only valid if $\hat{\lambda}_{min}(\cdot)$ is "well separated " from the other eigenvalues (see Appendix A). This -well separatedness - restriction is also always satisfied asymptotically in T, since it amounts to require (A 1) :

$$\lambda_i(\hat{H}(\cdot)) \overset{\min}{\neq} \hat{\lambda}_{min}(\cdot) \quad |\lambda_i(\hat{H}(\cdot)) - \hat{\lambda}_{min}(\cdot)| >> \frac{1}{\sqrt{T}} \tag{35}$$

with $\hat{\lambda}_{min}(\cdot) = O(1/\sqrt{T})$ and the other eigenvalues $\lambda_i(\hat{H}(\cdot))$ essentially constant. However the verification of this restriction can be checked on the numerical values of the computed eigenvalues and its non-respect detected and taken into account.

7 - SIMULATION RESULTS

In this section, we illustrate the behavior of the proposed test on some simulated examples.

Remember that our strategy consist in comparing for each matrix belonging to a sequence of matrices, a quantity μ defined in (33) to some threshold t (34) and to stop as soon as μ becomes -say- smaller than t. To compute μ, only the smallest, in absolute value, eigenvalue and its associated eigenvector (needed in the evaluation of $\hat{\sigma}^2_{min}(\cdot)$) of each matrix are required. Using then some results established in Appendix B, we introduce now a simplified way to obtain these quantities -"smallest eigenpair"-, thus further reducing the computations involved in the implementation of the presented test.

7.1 - Simplified computation of the "smallest eigenpair"

Under the assumption that the matrix $\hat{H}(\hat{p},\hat{q})$ admits a single and "well separated", small eigenvalue, it is proven in Appendix B that estimates (B1, B2) -denoted $\hat{\lambda}_{min,LS}(\cdot)$ and $\hat{X}_{LS}$- of $\hat{\lambda}_{min}(\hat{p},\hat{q})$ and an associated eigenvector $\hat{X}$ can be obtained from the solution X'_{LS} in the least-square (LS) sense of the following system of $\hat{p}$-linear equation in the $(\hat{p}-1)$-dimensional unknown vector X' :

$$\begin{bmatrix} \hat{r}_{\hat{q}-\hat{p}+1} & \hat{r}_{\hat{q}-\hat{p}+2} & \cdots & \hat{r}_{\hat{q}-1} \\ \hat{r}_{\hat{q}-\hat{p}+2} & & & \\ \vdots & & & \\ \hat{r}_{\hat{q}} & & & \hat{r}_{\hat{q}+\hat{p}-2} \end{bmatrix} X' = - \begin{bmatrix} \hat{r}_{\hat{q}} \\ \hat{r}_{\hat{q}+1} \\ \vdots \\ \hat{r}_{\hat{q}+\hat{p}-1} \end{bmatrix} \qquad (36)$$

This system is of the form $AX' = -b$ with A the $(\hat{p}-1)$ first columus of $\hat{H}(\hat{p},\hat{q})$ and b the last column of $\hat{H}(\hat{p},\hat{q})$. Let X'_{LS} be its least-square solution then

$$\hat{X}^T_{LS} = [\, X'^T_{LS} \vdots 1 \,]$$

and its Raylegh quotient for $\hat{H}^2(\cdot)$

$$\hat{\lambda}^2_{min,LS}(\hat{p},\hat{q}) = \hat{X}^T_{LS}\, \hat{H}^2(\hat{p},\hat{q})\, \hat{X}_{LS} \,/\, \hat{X}^T_{LS}\, \hat{X}_{LS} \qquad (37)$$

are estimates of the exact "smallest eigenpair" of $\hat{H}(\hat{p},\hat{q})$ coherent with the order of magnitudes we are considering. The reader can recognize in (36) a set of $\hat{p}$-consecutive Yule-Walker equations. As indicated in Appendix B, other estimates of the eigenpair can be obtained by only retaining the first $(\hat{p}-1)$-equations in (36). The corresponding estimates (B4, B5) will be denoted $\hat{\lambda}_{min,YW}(\cdot)$ and X_{YW} below.

Note again that these results hold under assumptions on the "smallest" eigenvalue, which translated in our context are identical to those required for the matrix perturbation results to be valid : the matrix $\hat{H}(\cdot)$ belongs to Class 1 or 2 (i.e. $\hat{\lambda}_{min}(\cdot) = 0\ (T^{-1/2})$) and T is large enough for $\hat{\lambda}_{min}(\cdot)$ to be well separated from the other eigenvalues (35). If these assumptions are note satisfied, the estimates obtained by these simplified computations can differ significantly from their exact values. Remember however that in this case, these true values are of limited interest since they are used in formulas (e.g. (30)) that are no longer valid. The above mentioned discrepancies are thus not troublesome except possibly in the case where the estimate of $\hat{\lambda}_{min}(\cdot)$ is small while its true value is quite high (i.e. $>> 0(T^{-1/2})$, as expected for matrices not belonging to Class 1 and 2). While this situations never

happens for $\hat{\lambda}_{min,LS}(\cdot)$ (37) (see (B 3)), it is not excluded for $\hat{\lambda}_{min,YW}(\cdot)$ and can lead to the false decision of declaring $\hat{H}(\cdot)$ singular. It is possible to detect this kind of situation by forming the residual vector :

$$r = \hat{H}(\cdot)\ \hat{\bar{X}}_{YW} - \hat{\lambda}_{min,YW}(\cdot)\ \hat{\bar{X}}_{YW}$$

where $\hat{\bar{X}}_{YW}$ is the normalized eigenvector. The (euclidean) norm of r - say $\delta = \|r\|$ - allows to test the accuracy of the approximate eigenpair $(\hat{\lambda}_{min,YW}(\cdot)\ ;\ \hat{X}_{YW})$. For an exact eigenpair δ is equal to zero and in general, for a symmetric matrix, the following result holds [19]

$$\min_i\ |\lambda_i(\hat{H}(\cdot)) - \hat{\lambda}_{min,YW}(\cdot)| \leqslant \delta\ .$$

It is thus possible to detect too large estimation errors and take corresponding decisions. We shall however not develop these possibilities and only present some simulation results using the true "smallest" eigenpair and the least-square estimated (37) smallest "eigenpair".

7.2 - Experimental results

Let us compare the results obtained with our approach to those obtained in [6] where the orders (p,q) of ARMA-processes are estimated by minimizing, with respect to p and q, the criterion :

$$\log \hat{\sigma}^2(p,q) + (p+q)\ \text{Log}\ T/T$$

with $\hat{\sigma}^2(p,q)$, the maximum likelihood estimate of the variance of the innovations. Note however that in [6], $\hat{\sigma}^2(p,q)$ is estimated using an economical procedure avoiding the "computationally troublesome maximum likelihood calculation".

As in [6] we only determine n, the maximum of p and q, i.e. we determine the order n of a model ARMA (n,n) which best represents the data. The threshold t in (34) has been taken equal to 3. There is no theoretical justification behind this choice which corresponds to an (asymptotic) probability of false alarm of 9 %. The results, we have obtained are summarized in Table 1, where the first column of each row should be read as follows : (true AR order p ; true MA order q ; AR coefficients $a_1, a_2, \ldots, a_p$; MA coefficients, $c_1, c_2, \ldots, c_q$; the number of data T, the number of independant trials). The remaining triplets indicate the number of times the corresponding order has been obtained respectively with our approach, the simplified LS version of our approach and the approach in [6].

The performances of the three methods are quite similar, so that as far as our approach is concerned, the simplified LS version should be prefered. In row 6 however our results are worse than those of [6]. In [6], the results of rows 5 and 6 were described as bad and the authors explain that the corresponding AR2 process is almost perfectly predictable with an ARMA (1,1) model. In our case, the explanation is different, the exact full rank matrix H(2,2) possesses an eigenvalue $\lambda_{min}(2,2) = -4.09 \cdot 10^{-2}$ which is quite small and $\hat{H}(2,2)$ will in general be declared singular for small sample sizes T (under exact covariance information, equality in (30) is achieved with t = 3 for T = 1314, so that the results obtained for T = 100 and 500 are surprisingly good).

Model parameters	n = 0	n = 1	n = 2	n = 3	n = 4	n = 5
1, 1, -0.5, 0.8, 100, 50		47, 47, 46	3, 3, 4			
2, 1, 0.64 0.7, 0.8, 25, 30		3, 2, 5	25, 26, 23	2, 1, 1	0, 1, 0	0, 0, 1
——— " ——— 50, 30			28, 28, 27	2, 2, 3	0, 1, 0	
——— " ——— 100, 30			28, 28, 29	2, 1, 1	0, 1, 0	0, 0, 1
2, 0, -1.1, 0.24, 100, 30		26, 26, 26	4, 4, 4			
——— " ———, 500, 30		16, 20, 12	12, 10, 18	2, 0, 0		
1, 1, 0 3, 0 5, 500, 30	0, 0, 2	28, 27, 27	2, 3, 0			0, 0, 1

TABLE 1 : Distribution of estimated order, respectively (the proposed approach, the simplified version, the approach in [6])

8 - CONCLUDING REMARKS AND EXTENSIONS

The proposed procedure for determining the orders of an ARMA-process provides an alternative to the usual criteria based on maximum likelihood, which from a practical point of view involve a much higher computational burden.

One can think of different extensions to the proposed approach. There is no real justification for preferring the Hankel matrices chosen here to the Toeplitz matrices generally considered in this kind of approaches, except possibly the symmetry which leads to simpler formulas, real eigenvalues and in general to well conditionned algorithms. Of course, any family of matrices whose rank is ARMA-order dependent can be considered. We have performed a number of simulations for different such families. A promising family, we considered at length [21], is constituted by the matrices - $\bar{H}_k$, $k = 1,2,...$ - the singular values of which converge towards the canonical correlation coefficients of the process [22,23]

$$\bar{H}_k = L_k^{-1} H(k,k) L_k^{-T}$$

with L_k the Cholesky factor of $R(k,0)$. It is straightforward to tailor the results of this paper so as to obtain the corresponding test. While the computations for the $\bar{H}_k$ family are more involved than those considered here for the $H(p,q)$ family, the experimental results appear to be quite similar. The same comments apply to the matrices

$$H(k,k)\ R(k,0)^{-1}\ H(k,k), \qquad k = 1,2,...$$

advocated for in [24].

Within a family of matrices, it seems natural to investigate a test for several zero eigenvalues. When comparing our approach to the likelihood approaches, it appears that we use very little information (the beginning of the covariance sequence) in contrast to these methods. Performing the singular value decomposition of -say- $\hat{H}(N,N)$ for quite large N and testing for the number of zero singular values would allow to use more information. Unfortunatly we do not know how to generalize our test to the multiple zero-eigenvalue case in a simple fashion. Another mean to increase the amount of information taken into account consists in considering rectangular matrices, the corresponding test is under investigation.

APPENDIX A : A MATRIX PERTURBATION RESULT

The purpose of this appendix is to present a simple proof of the result from matrix perturbation theory we use in the text. The proof we present is derived from [17] and can be easily extended to obtain more general results.

Let A and B be real symmetric matrices. Since eigenvalues are continuous functions of the matrix components, it makes sense to investigate by how much a simple eigenvalue λ of A differs from the corresponding eigenvalue in $A+\varepsilon B$ for small εB, i.e. for small ε and (spectral) norm of B of the order of $\varepsilon^\circ = 1$, $\|B\|_2 = O(1)$. We assume further λ to be "well separated" from the other eigenvalues of A :

$$\min_{\lambda_i(A) \neq \lambda} |\lambda - \lambda_i(A)| = O(1) >> \varepsilon \qquad (A1)$$

Let X be the normalized eigenvector, taken real, corresponding to λ. We way now construct an orthogonal matrix P with first column X, such that :

$$\Lambda = P^T A P = \begin{bmatrix} \lambda & 0 \\ 0 & \Lambda_1 \end{bmatrix} \qquad (A2)$$

(one can take as P the orthogonal matrix which diagonalizes A, e.g.).

The structure of Λ, see the last matrix in (A2), clearly isolates λ and allows to rewritte (A1) as :

$$\min_{\lambda_i(\Lambda_1)} |\lambda - \lambda_i(\Lambda_1)| = O(1) \qquad (A3)$$

Applying the same congruence transformation to the perturbed matrix $A+\varepsilon B$ yields :

$$P^T(A+\varepsilon B)P = \Lambda + \varepsilon P^T B\, P = \Lambda + \varepsilon D \tag{A4}$$

a matrix that no longer possesses the structure in (A2) characteristic of an eigenvalue equal to λ. Let us partition $D = P^T B\, P$

$$D = \begin{bmatrix} d & d_1^T \\ d_1 & D_1 \end{bmatrix}$$

with d a scalar equal to $X^T B X$ and establish that the perturbed matrix possesses an eigenvalue $\hat{\lambda}$ satisfying :

$$\begin{aligned} \hat{\lambda} &= \lambda + \varepsilon d + O(\varepsilon^2) \\ &= \lambda + \varepsilon X^T B X + O(\varepsilon^2) \end{aligned} \tag{A5}$$

We merely verify that there exists a further congruence transformation with orthogonal matrix Q, such that :

$$Q^T(\Lambda + \varepsilon D)Q = \begin{bmatrix} \lambda + \varepsilon d & 0 \\ 0 & \Lambda_1 + \varepsilon D_1 \end{bmatrix} + O(\varepsilon^2) \tag{A6}$$

this will establish (A5). Let us consider a candidate orthogonal matrix of the form

$$Q = I + \varepsilon U + O(\varepsilon^2)$$

It has to satisfy $Q^{-1} = Q^T$ and since for small εU :

$$Q^{-1} = I - \varepsilon U + O(\varepsilon^2) \tag{A7}$$

this implies $U^T = -U$. Partitionning U as in (A2), we choose to take it equal to :

$$U = \begin{bmatrix} 0 & -u_1^T \\ u_1 & 0 \end{bmatrix}$$

with u_1 a column vector yet to be specified. We then have :

$$Q^T(\Lambda + \varepsilon D)Q = \Lambda + \varepsilon \begin{bmatrix} d & d_1^T - u_1^T(\lambda I - \Lambda_1) \\ d_1 - (\lambda I - \Lambda_1)u_1 & D_1 \end{bmatrix} + O(\varepsilon^2)$$

which, if we decide to take :

$$u_1 = (\lambda I - \Lambda_1)^{-1} d_1 \qquad \text{(A8)}$$

leads to relation (A6). Premultiply now (A6) by PQ and combine it with (A4) :

$$(A + \varepsilon B)PQ = PQ \left\{ \begin{bmatrix} \lambda + \varepsilon d & 0 \\ 0 & \Lambda_1 + \varepsilon D_1 \end{bmatrix} + O(\varepsilon^2) \right\}$$

to establish (A5) and the following property of $\hat{X}$, the perturbed eigenvector, by equating the first columns :

$$(A + \varepsilon B)\hat{X} = (\lambda + \varepsilon d + O(\varepsilon^2))\hat{X}$$

$$\hat{X} = X + O(\varepsilon) \qquad \text{(A9)}$$

Note that assumption (A1, A3) is required to guarantee that u_1 in (A8) is of order 1 and thus εU of order ε and (A7) valid.

APPENDIX B

Proposition : Let B be a real symmetric matrix with orthogonal eigen-decomposition

$$B = \sum_{1}^{n} \lambda_i u_i u_i^T$$

where :

$$\lambda_1 = O(\varepsilon)$$

$$\min_{i \neq 1} |\lambda_i - \lambda_1| = O(1) \qquad (>> \varepsilon)$$

and the last component of the eigenvector u_1, $u_1(n)$ is O(1). Then partitioning the matrix B as :

$$B = \left[C \;\vdots\; d \right]$$

with d a column-vector and denoting Z' the least square solution to CZ' = -d the following results hold :

$$Z \triangleq \begin{bmatrix} Z' \\ 1 \end{bmatrix} = \frac{1}{u_1(n)} u_1 + O(\varepsilon^2) \qquad \text{(B1)}$$

$$\frac{Z^T B^2 Z}{Z^T Z} = \lambda_1^2 + O(\varepsilon^4) \qquad \text{(B2)}$$

Proof : Consider the optimization problem

$$\min X^T B^2 X$$

subject to

$$X^T e_n = 1$$

with $$e_n^T = [0 \ldots 0 \ 1] \quad .$$

One verifies that the extremum is reached for X = Z as defined in (B1). Now from elementary optimization theory [20], the solution Z also satisfies :

$$B^2 Z = \mu \, e_n$$

$$Z^T e_n = 1$$

using then the eigendecomposition of B one has

$$Z = \mu \sum \frac{1}{\lambda_i^2} \, u_i(n) \, u_i$$

and imposing $Z^T e_n = 1$ yields

$$\mu = \left(\sum \frac{u_i(n)^2}{\lambda_i^2} \right)^{-1}$$

hence

$$Z = \left(\sum \frac{u_i(n)^2}{\lambda_i^2} \right)^{-1} \sum \frac{u_i(n)}{\lambda_i^2} \, u_i$$

and after some manipulations

$$\frac{Z^T B^2 Z}{Z^T Z} = \left(\sum \frac{u_i(n)^2}{\lambda_i^4} \right)^{-1} \left(\sum \frac{u_i(n)^2}{\lambda_i^2} \right)$$

The announced results then follows easily noting e.g. that :

$$\sum \frac{u_i(n)^2}{\lambda_i^2} = \frac{u_1(n)^2}{\lambda_1^2} + O(1) = \frac{u_1(n)^2 + \lambda_1^2 \, O(1)}{\lambda_1^2} = \frac{u_1(n)^2 + O(\varepsilon^2)}{\lambda_1^2}$$

and analoguous derivations. One can also verify that, it any case

$$\frac{z^T B^2 z}{z^T z} \geq \lambda_1^2 \qquad \text{(B3)}$$

since this relation can be rewritten :

$$\sum \frac{\lambda_1^2}{\lambda_i^2} u_i^2(n) \geq \sum \frac{\lambda_1^4}{\lambda_i^4} u_i(n)^2$$

where

$$\left(\frac{\lambda_1}{\lambda_i}\right)^2 \leq 1 \qquad \forall i$$

■

This proposition says that if a symmetric matrix B possesses a single and well isolated eigenvalue close to zero, then this eigenvalue can be estimated with specified precision by solving an overdetermined system in a least square sense. A difficulty appears if $u_1(n)$, which is unknown a priori is itself small, but then the matrix $B^T B$ will be almost singular, an easily detected situation. Another proposition concerning the same problem is the following.

Proposition : Let B be a real symmetric matrix with orthogonal eigendecomposition

$$B = \sum_{i=1}^{n} \lambda_i u_i u_i^T$$

where

$$\lambda_1 = O(\varepsilon)$$

$$\min_{i \neq 1} |\lambda_i - \lambda_1| = O(1) \qquad (>> \varepsilon)$$

and the last component of the eigenvector u_1, $u_1(n)$ is O(1). Then partitionning B as

$$B = \begin{bmatrix} A & b \\ b^T & d \end{bmatrix}$$

with d a scalar and setting $Y' = -A^{-1}b$ the following results hold :

$$Y \triangleq \begin{bmatrix} Y' \\ 1 \end{bmatrix} = \frac{1}{u_1(n)} u_1 + O(\varepsilon) \qquad \text{(B4)}$$

$$\frac{Y^T B Y}{Y^T Y} = \lambda_1 + O(\varepsilon^2) \qquad \text{(B5)}$$

▲

The proof is analoguous, once one has noted that Y (B4) can be seen as part of the solution to

$$B\,Y = \mu\, e_n$$

$$Y^T e_n = 1$$

a linear system in the unknowns Y (a vector) and μ (a scalar).

REFERENCES

[1] AKAIKE H. "Fitting autoregressive models for prediction", Ann. Inst. Statist. Math., vol 21, pp. 243-247, 1969.

[2] AKAIKE H. " A new look at the statistical model identification", IEEE-Trans. Automat. Control, vol AC-19, pp. 716-723, Dec. 1974.

[3] SCHWARZ G. : "Estimating the dimension of a model", Ann. Statist., vol 6, pp 461-464, 1978.

[4] RISSANEN J. : "Modeling by shortest data description", Automatica, vol 14, pp. 467-471, 1978.

[5] RISSANEN J. : "A universal prior for integers and estimation by minimum description length", Ann. Statist., vol 11, pp. 416-431, 1983.

[6] HANNAN E.J., RISSANEN J. : "Recursive estimation of mixed ARMA-order", Biometrika, vol 69, pp. 81-94, 1982.

[7] LEE D.T.L. : "System order estimation of ARMA models by ladder canonical forms", Proc. ACC, pp. 1002-1007, 1983.

[8] GANTMACHER F. : The theory of Matrices, Chelsea, New York, 1959.

[9] STOICA P. : "On a procedure for testing the orders of time series", IEEE-Trans. Autom. Control, vol 26, pp. 572-573, 1981.

[10] STEWART G.W. : "Rank degeneracy", SIAM J. Sc. Stat. Comput., vol 5, pp. 403-413, 1984.

[11] CHOW J.C. : "On estimating the orders of on ARMA process with uncertain observations", IEEE-Trans. Autom. Control, vol 17, pp. 707-709, 1972.

[12] CHAN Y.T., WOOD J.C. : "A new order determination technique for ARMA processes", IEEE-Trans. ASSP, vol 32, pp. 517-521, 1984.

[13] HANNAN E.J., HEYDE C.C. : "On limit theorems for quadratic functions of discrete time series", Ann. of Math. Statistics, vol 43, pp. 2058-2066, 1972.

[14] WILKINSON J.H. : The Algebraic Eigenvalue Problem, Oxford Univ. Press, Oxford, 1965.

[15] ANDERSON T.W. : The Statistical Analysis of Time Series, New York, Wiley, 1971.

[16] KATO T. : Pertubation Theory for Linear Operators, Berlin, Springer, 1976.

[17] VOIEVODINE V. : Principes numériques d'algèbre linéaire, MOSCOU, Editions Mir, 1980.

[18] ANDERSON O.D. : "The recursive nature of the stationarity and invertibility restraints of the parameters of ARMA processes", Biometrika, vol 62, n° 3, pp. 706-707, 1975.

[19] ORTEGA J.M. : Numerical Analysis, Academic Press, New York, 1972.

[20] LUENBERGER D.G. : Introduction to linear and non linear Programming, Addison-Wesley, Reading, 1973.

[21] FUCHS J.J. : "ARMA-order estimation via canonical correlation analysis", submitted to C.D.C. 1985 and rejected.

[22] JEWELL N.P., BLOOMFIELD P. : "Canonical correlation of past and future for time-series", Annals of Statistics, vol 11, pp. 837-855, 1983.

[23] DESAI U.B., PAL B. : "A realization approach to stochastic model reduction and balanced stochastic realization", Proc. 21st C.D.C., pp. 1105-1112, 1982.

[24] ARUN S.K. : "A principal component approach to approximate modeling and ARMA spectral estimation", Ph. D. Dissertation, E.E. Dept., USC, L.A., 1984.